

METHOD AND APPARATUS FOR MULTIVARIATE SPACE PROCESSING

BACKGROUND OF THE INVENTION

5 1. Field of the Invention

The present invention relates to a multivariate space processing technology and multivariate space processing method and apparatus, and more particularly relates to a method and apparatus for predicting, interpolating and displaying 10 multivariate data.

2. Description of the Related Art

Scientific visualization is often used as a tool to allow a user to intuitively grasp or understand complex 15 phenomena and large-scale data. For example, in order to visualize time-varying complex three-dimensional phenomena such as ocean currents or tornados, density and other attributes of each point in the three-dimensional space are assigned for each voxel and time-varied values are computed 20 sequentially and displayed, so that a general structure and motion of the large-scale data can be approximated. Besides grasping a phenomenon as a whole, visualization is also used as a tool to gauge the effectiveness of a hypothetical modeling simulation in a simplified and convenient manner.

In most cases, scientific visualization is implemented to allow a phenomenon or model to be understood in a general and simplified way. However, in actuality, the creation of a 5 scientific visualization will generally involve a series of relatively complex processes such as setting up complex equations and functions, tracing time-varying values of each point in the n-dimensional space, and converting this large amount of data to a display space. Thus, the computational 10 load for even a basic scientific visualization can be quite large, so that, in many cases, the amount of time and effort required to complete the scientific visualization adversely affects the pace at which the overall research processes can be conducted.

15

SUMMARY OF THE INVENTION

The present invention has been made in view of the 20 foregoing circumstances and an object thereof is to provide a technique by which the time-varying behaviour of three-dimensional objects can be visualized or displayed with a reduced computational load. Another object of the present invention is to provide a technique by which the structure of

or changes in a multivariate space having a dimension greater than three can be visualized or displayed in a simple and efficient manner.

Hereinafter, the term "variate" is used to describe each 5 of a plurality of variable quantities in a multivariate space. A variate may also sometimes be called a parameter and these terms may generally be treated in the same manner as the term "dimension" as the case may be. Further, when a variate is referred to as a variate of an object, the variate may also be 10 defined as an attribute of the object.

An embodiment according to the present invention relates to a multivariate space processing method. The multivariate space processing method includes: degenerating multivariate data into three predetermined variates (referred to as, for 15 example, x, y, t hereinafter); determining a reference variate (referred to as, for example, t hereinafter) to serve as a reference among the three variates; acquiring a first two-dimensional space formed by the remaining two variates when the reference variate takes a first value; acquiring a second 20 two-dimensional space formed by the remaining two variates when the reference variate takes a second value; and computing a matching between the first two-dimensional space and the second two-dimensional space.

In particular, the "degenerating" may be such that

certain variate(s) or dimension(s) are simply cut off or, in other cases, may be such that a specific value is assumed for certain variate(s) or dimensions. For example, in a case where a three-dimensional space is degenerated to a two-

5 dimensional space, the variate of height may be set to some constant ($z=const.$). Here, it suffices that the degeneration is such that an n dimension space is degenerated to a dimension which is less than n . For example, a three-dimensional space may simply be converted to an arbitrary 10 plane $ax+by+cz+d=0$. Moreover, a three-dimensional space may be converted to a curved surface such as $ax^2+by+c=0$. In any event, reducing the number of variates is called degeneration.

As an example, consider the case where the reference variate t is time. Now, the remaining two variates x and y 15 can be used to define a first two-dimensional space, sometimes called the first image, at time $t=t_0$ and a second two-dimensional space, sometimes called the second image, at time $t=t_1$. If an image matching is computed between the first and second images, the approximate behavior of the two variates x 20 and y in the region $t=[t_0, t_1]$ can be determined. In this case, it suffices to compute a matching between the two-dimensional images, and the computational load can be reduced by selecting a suitable algorithm. The "base technology" described later provides one such suitable algorithm and

contributes to producing a suitable visualization.

With this embodiment, a complicated multivariate space or object can be converted into a simple model by selecting three variates. As an example, consider that there are a 5 large number of factors (variates) that go into determining a stock price. Initially, three variates of the stock price are tentatively determined based on a rule of thumb or an arbitrary selection after a search algorithm has enumerated a number of possible variates. Next, a first image and a second 10 image are generated and an intermediate image thereof at, for example, $t=t_2$ is generated by matching and interpolation. Thereafter, this intermediate image (also referred to as a virtual intermediate image hereinafter) and the actual stock price at the time $t=t_2$ (also referred to as an authentic 15 intermediate image) are compared to determine the suitability of the selection of the three variates. Thus, if the virtual intermediate image is close to the authentic intermediate image, it is concluded that the three variates have high importance. This can be used or further tested by determining 20 if it is possible to analyze or predict the stock price at another time by using these three variates. Further, the virtual intermediate image and the authentic intermediate image may also be compared while the three variates are being changed, while the selection of the variate serving as the

reference (also referred to as the reference variate) is being changed, and/or while the value of the reference variate is being changed.

By adopting the above-described methods while reducing 5 the number of parameters to three, a stock price model can be simultaneously simplified and optimized. The methods may be stated from a more general perspective as: viewing a multidimensional phenomenon or object from various angles while reducing parameters one at a time, conducting an 10 inspection to find an angle at which a most characteristic image is obtained, and repeating a degeneration to view the object at that particular angle. Using a simple example, consider a case where a three-dimensional object is projected as a "shadow picture" on a plane. There will generally be an 15 angle at which the original object can be relatively easily grasped or understood from the shadow picture. A projection operation using this angle corresponds to the degeneration.

The above-described matching may be performed by regarding the first two-dimensional space and second two- 20 dimensional space as a first image and a second image, respectively, and computing pixel by pixel based on correspondence between a critical point detected through a two-dimensional search on the first image and a critical point detected through a two-dimensional search on the second image.

- The matching may further include: multiresolutionalizing the first image and the second image by respectively extracting the critical points; performing a pixel-by-pixel matching computation on the first image and the second image
- 5 at the same multiresolution level; and acquiring a pixel-by-pixel correspondence relation at a finer level of resolution while inheriting a result of the pixel-by-pixel matching computation from a matching computation at a different multiresolution level.
- 10 Another preferred embodiment according to the present invention relates also to a multivariate space processing method that includes: acquiring a first image and a second image by projecting three-dimensional data on a predetermined plane; and computing a matching between the first image and the second image. By implementing this method, after the three-dimensional data of, for example, a tornado are projected on a two-dimensional plane to create the first and second images, intermediate states in a region of time $t=[t_0, t_1]$ can be visualized by interpolating the two-dimensional
- 15 images for time $t=t_0$ and time $t=t_1$. Computationally, once the matching is completed, there will not be much time needed for the interpolation. Moreover, since the interpolation can be performed at any point in the region by arbitrarily varying the time t , the computational load becomes extremely low

compared to conventional visualization, in which a three-dimensional computation is performed sequentially. Thus, this embodiment is effective particularly when one wants to quickly visualize a phenomenon or simulation in as simplified and 5 convenient a manner as possible.

Still another embodiment according to the present invention relates to a multivariate space processing apparatus that includes: a preprocessing unit which degenerates multivariate data into three predetermined variates; a 10 conversion unit which determines a reference variate from among the three variates to serve as a reference, acquires, as a first image, a two-dimensional space formed by the remaining two variates when the reference variate takes a first value, and acquires, as a second image, a two-dimensional space 15 formed by the remaining two variates when the reference variate takes a second value; and a matching processor which computes a matching between the first image and the second image. In this embodiment, the apparatus may further include an intermediate image generator which generates an 20 intermediate image of the first image and the second image by performing an interpolation computation based on a result of the matching computation.

Still another embodiment according to the present invention relates also to a multivariate space processing

apparatus that includes: a conversion unit which acquires a first image and a second image by projecting three-dimensional data on a predetermined plane; and a matching processor which computes a matching between the first image and the second 5 image.

In the embodiments herein, the matching and, in particular, the matching using critical points may be an application of the techniques (referred to as the "base technology" hereinafter) proposed in Japanese Patent 10 No.2927350 owned by the same assignee of the present patent application.

It is to be noted that the base technology is not a prerequisite in the present invention. Moreover, it is also possible to have replacement or substitution of the above- 15 described components, elements, functions or processes in part or whole as between method and apparatus or to add components, elements, functions or processes to method or apparatus, and it will also be understood that the components, elements, 20 functions or processes may be implemented by a computer program and saved on a recording medium or the like and are all effective as and encompassed by the present invention.

Moreover, this summary of the invention does not necessarily describe all necessary features so that the invention may also be sub-combination of these described

features.

BRIEF DESCRIPTION OF THE DRAWINGS

5

Fig. 1(a) is an image obtained as a result of the application of an averaging filter to a human facial image.

Fig. 1(b) is an image obtained as a result of the application of an averaging filter to another human facial 10 image.

Fig. 1(c) is an image of a human face at $p^{(5,0)}$ obtained in a preferred embodiment in the base technology.

Fig. 1(d) is another image of a human face at $p^{(5,0)}$ obtained in a preferred embodiment in the base technology.

15 Fig. 1(e) is an image of a human face at $p^{(5,1)}$ obtained in a preferred embodiment in the base technology.

Fig. 1(f) is another image of a human face at $p^{(5,1)}$ obtained in a preferred embodiment in the base technology.

20 Fig. 1(g) is an image of a human face at $p^{(5,2)}$ obtained in a preferred embodiment in the base technology.

Fig. 1(h) is another image of a human face at $p^{(5,2)}$ obtained in a preferred embodiment in the base technology.

Fig. 1(i) is an image of a human face at $p^{(5,3)}$ obtained in a preferred embodiment in the base technology.

Fig. 1(j) is another image of a human face at $p^{(5,3)}$ obtained in a preferred embodiment in the base technology.

Fig. 2(R) shows an original quadrilateral.

Fig. 2(A) shows an inherited quadrilateral.

5 Fig. 2(B) shows an inherited quadrilateral.

Fig. 2(C) shows an inherited quadrilateral.

Fig. 2(D) shows an inherited quadrilateral.

Fig. 2(E) shows an inherited quadrilateral.

10 Fig. 3 is a diagram showing the relationship between a source image and a destination image and that between the m -th level and the $(m-1)$ th level, using a quadrilateral.

Fig. 4 shows the relationship between a parameter η (represented by x-axis) and energy C_f (represented by y-axis).

15 Fig. 5(a) is a diagram illustrating determination of whether or not the mapping for a certain point satisfies the bijectivity condition through the outer product computation.

Fig. 5(b) is a diagram illustrating determination of whether or not the mapping for a certain point satisfies the bijectivity condition through the outer product computation.

20 Fig. 6 is a flowchart of the entire procedure of a preferred embodiment in the base technology.

Fig. 7 is a flowchart showing the details of the process at S1 in Fig. 6.

Fig. 8 is a flowchart showing the details of the process

at S10 in Fig. 7.

Fig. 9 is a diagram showing correspondence between partial images of the m -th and $(m-1)$ -th levels of resolution.

Fig. 10 is a diagram showing source hierarchical images 5 generated in the embodiment in the base technology.

Fig. 11 is a flowchart of a preparation procedure for S2 in Fig. 6.

Fig. 12 is a flowchart showing the details of the process at S2 in Fig. 6.

10 Fig. 13 is a diagram showing the way a submapping is determined at the 0-th level.

Fig. 14 is a diagram showing the way a submapping is determined at the first level.

15 Fig. 15 is a flowchart showing the details of the process at S21 in Fig. 12.

Fig. 16 is a graph showing the behavior of energy $C_f^{(m,s)}$ corresponding to $f^{(m,s)}$ ($\lambda = i\Delta\lambda$) which has been obtained for a certain $f^{(m,s)}$ while varying λ .

20 Fig. 17 is a diagram showing the behavior of energy $C_f^{(n)}$ corresponding to $f^{(n)}$ ($\eta = i\Delta\eta$) ($i = 0, 1, \dots$) which has been obtained while varying η .

Fig. 18 shows a structure of a multivariate space processing apparatus according to an embodiment of the

invention.

Fig. 19 is a flowchart showing a multivariate space processing method according to an embodiment of the invention.

Fig. 20 shows more detail of S300 shown in Fig. 19.

5 Fig. 21 shows more detail of S302 shown in Fig. 19.

Fig. 22 shows more detail of S308 shown in Fig. 19.

Fig. 23 shows an image based on data for a tornado at time $t=t_1$.

Fig. 24 shows an image based on data for a tornado at

10 time $t=t_2$.

Fig. 25 shows an intermediate image obtained based on the images of Figs. 23 and 24.

15 DETAILED DESCRIPTION OF THE INVENTION

The invention will now be described based on the preferred embodiments, which are not intended to limit the scope but to exemplify the present invention. All of the 20 features and the combinations thereof described in the embodiments are not necessarily essential to the invention.

First, the multiresolutional critical point filter technology and the image matching processing using the technology, both of which will be utilized in the preferred

embodiments, will be described in detail as "Base Technology".

Namely, the following sections [1] and [2] (below) belong to the base technology, where section [1] describes elemental techniques and section [2] describes a processing procedure.

- 5 These techniques are patented under Japanese Patent No. 2927350 and owned by the same assignee of the present invention. As described in more detail below following the discussion of the base technology, according to the
- embodiments of the present invention there is provided a mesh
- 10 on an image, so that lattice points of the mesh represent a plurality of pixels of the image. Thus, even though application efficiency for a pixel-by-pixel matching technique as described in the base technology is naturally high, it is to be noted that the image matching techniques provided in the
- 15 present embodiments are not limited to the same levels.

After discussing the base technology, there is a description, related to Figs. 18 to 23, of image data coding and decoding techniques for multivariate space processing according to embodiments of the present invention and

20 utilizing the base technology.

Base Technology

[1] Detailed description of elemental techniques

[1.1] Introduction

Using a set of new multiresolutional filters called critical point filters, image matching is accurately computed. There is no need for any prior knowledge concerning the content of the images or objects in question. The matching of 5 the images is computed at each resolution while proceeding through the resolution hierarchy. The resolution hierarchy proceeds from a coarse level to a fine level. Parameters necessary for the computation are set completely automatically by dynamical computation analogous to human visual systems. 10 Thus, There is no need to manually specify the correspondence of points between the images.

The base technology can be applied to, for instance, completely automated morphing, object recognition, stereo photogrammetry, volume rendering, and smooth generation of 15 motion images from a small number of frames. When applied to morphing, given images can be automatically transformed. When applied to volume rendering, intermediate images between cross sections can be accurately reconstructed, even when a distance between cross sections is rather large and the cross sections 20 vary widely in shape.

[1.2] The hierarchy of the critical point filters

The multiresolutional filters according to the base technology preserve the intensity and location of each

critical point included in the images while reducing the resolution. Initially, let the width of an image to be examined be N and the height of the image be M . For simplicity, assume that $N=M=2n$ where n is a positive integer.

- 5 An interval $[0, N] \subset \mathbb{R}$ is denoted by I . A pixel of the image at position (i, j) is denoted by $p^{(i,j)}$ where $i, j \in I$.

Here, a multiresolutional hierarchy is introduced.

Hierarchized image groups are produced by a multiresolutional filter. The multiresolutional filter carries out a two

- 10 dimensional search on an original image and detects critical points therefrom. The multiresolutional filter then extracts the critical points from the original image to construct another image having a lower resolution. Here, the size of each of the respective images of the m -th level is denoted as
- 15 $2^m \times 2^m$ ($0 < m < n$). A critical point filter constructs the following four new hierarchical images recursively, in the direction descending from n .

$$\begin{aligned} p_{(i,j)}^{(m,0)} &= \min(\min(p_{(2i,2j)}^{(m+1,0)}, p_{(2i,2j+1)}^{(m+1,0)}), \min(p_{(2i+1,2j)}^{(m+1,0)}, p_{(2i+1,2j+1)}^{(m+1,0)})) \\ p_{(i,j)}^{(m,1)} &= \max(\min(p_{(2i,2j)}^{(m+1,1)}, p_{(2i,2j+1)}^{(m+1,1)}), \min(p_{(2i+1,2j)}^{(m+1,1)}, p_{(2i+1,2j+1)}^{(m+1,1)})) \\ p_{(i,j)}^{(m,2)} &= \min(\max(p_{(2i,2j)}^{(m+1,2)}, p_{(2i,2j+1)}^{(m+1,2)}), \max(p_{(2i+1,2j)}^{(m+1,2)}, p_{(2i+1,2j+1)}^{(m+1,2)})) \\ p_{(i,j)}^{(m,3)} &= \max(\max(p_{(2i,2j)}^{(m+1,3)}, p_{(2i,2j+1)}^{(m+1,3)}), \max(p_{(2i+1,2j)}^{(m+1,3)}, p_{(2i+1,2j+1)}^{(m+1,3)})) \end{aligned}$$

--- (1)

- 20 where we let

$$p_{(i,j)}^{(n,0)} = p_{(i,j)}^{(n,1)} = p_{(i,j)}^{(n,2)} = p_{(i,j)}^{(n,3)} = p_{(i,j)} \quad \text{--- (2)}$$

The above four images are referred to as subimages hereinafter. When $\min_{x \leq t \leq x+1}$ and $\max_{x \leq t \leq x+1}$ are abbreviated to α and β , respectively, the subimages can be expressed as 5 follows:

$$\begin{aligned} P^{(m,0)} &= \alpha(x)\alpha(y)p^{(m+1,0)} \\ P^{(m,1)} &= \alpha(x)\beta(y)p^{(m+1,1)} \\ P^{(m,2)} &= \beta(x)\alpha(y)p^{(m+1,2)} \\ P^{(m,3)} &= \beta(x)\beta(y)p^{(m+1,3)} \end{aligned}$$

Namely, they can be considered analogous to the tensor products of α and β . The subimages correspond to the respective critical points. As is apparent from the above 10 equations, the critical point filter detects a critical point of the original image for every block consisting of 2 X 2 pixels. In this detection, a point having a maximum pixel value and a point having a minimum pixel value are searched with respect to two directions, namely, vertical and 15 horizontal directions, in each block. Although pixel intensity is used as a pixel value in this base technology, various other values relating to the image may be used. A pixel having the maximum pixel values for the two directions, one having minimum pixel values for the two directions, and 20 one having a minimum pixel value for one direction and a maximum pixel value for the other direction are detected as a local maximum point, a local minimum point, and a saddle.

point, respectively.

By using the critical point filter, an image (1 pixel here) of a critical point detected inside each of the respective blocks serves to represent its block image (4 pixels here) in the next lower resolution level. Thus, the resolution of the image is reduced. From a singularity theoretical point of view, $\alpha(x)\alpha(y)$ preserves the local minimum point (minima point), $\beta(x)\beta(y)$ preserves the local maximum point (maxima point), $\alpha(x)\beta(y)$ and $\beta(x)\alpha(y)$ preserve the saddle points.

At the beginning, a critical point filtering process is applied separately to a source image and a destination image which are to be matching-computed. Thus, a series of image groups, namely, source hierarchical images and destination hierarchical images are generated. Four source hierarchical images and four destination hierarchical images are generated corresponding to the types of the critical points.

Thereafter, the source hierarchical images and the destination hierarchical images are matched in a series of resolution levels. First, the minima points are matched using $p^{(m,0)}$. Next, the first saddle points are matched using $p^{(m,1)}$ based on the previous matching result for the minima points. The second saddle points are matched using $p^{(m,2)}$. Finally, the maxima points are matched using $p^{(m,3)}$.

Figs. 1c and 1d show the subimages $p^{(5,0)}$ of the images in Figs. 1a and 1b, respectively. Similarly, Figs. 1e and 1f show the subimages $p^{(5,1)}$, Figs. 1g and 1h show the subimages $p^{(5,2)}$, and Figs. 1i and 1j show the subimages $p^{(5,3)}$.

5 Characteristic parts in the images can be easily matched using subimages. The eyes can be matched by $p^{(5,0)}$ since the eyes are the minima points of pixel intensity in a face. The mouths can be matched by $p^{(5,1)}$ since the mouths have low intensity in the horizontal direction. Vertical lines on both 10 sides of the necks become clear by $p^{(5,2)}$. The ears and bright parts of the cheeks become clear by $p^{(5,3)}$ since these are the maxima points of pixel intensity.

As described above, the characteristics of an image can be extracted by the critical point filter. Thus, by 15 comparing, for example, the characteristics of an image shot by a camera with the characteristics of several objects recorded in advance, an object shot by the camera can be identified.

20 [1.3] Computation of mapping between images

Now, for matching images, a pixel of the source image at the location (i,j) is denoted by $p_{(i,j)}^{(n)}$ and that of the destination image at (k,l) is denoted by $q_{(k,l)}^{(n)}$ where $i, j, k, l \in I$. The energy of the mapping between the images (described

later in more detail) is then defined. This energy is determined by the difference in the intensity of the pixel of the source image and its corresponding pixel of the destination image and the smoothness of the mapping. First,

5 the mapping $f^{(m,0)}: p^{(m,0)} \rightarrow q^{(m,0)}$ between $p^{(m,0)}$ and $q^{(m,0)}$ with the minimum energy is computed. Based on $f^{(m,0)}$, the mapping $f^{(m,1)}$ between $p^{(m,1)}$ and $q^{(m,1)}$ with the minimum energy is computed.

This process continues until $f^{(m,3)}$ between $p^{(m,3)}$ and $q^{(m,3)}$ is

computed. Each $f^{(m,i)}$ ($i = 0, 1, 2, \dots$) is referred to as a

10 submapping. The order of i will be rearranged as shown in the following equation (3) in computing $f^{(m,i)}$ for reasons to be described later.

$$f^{(m,i)}: p^{(m,\sigma(i))} \rightarrow q^{(m,\sigma(i))} \quad \text{--- (3)}$$

where $\sigma(i) \in \{0, 1, 2, 3\}$.

15

[1. 3. 1] Bijectivity

When the matching between a source image and a destination image is expressed by means of a mapping, that mapping shall satisfy the Bijectivity Conditions (BC) between

20 the two images (note that a one-to-one surjective mapping is called a bijection). This is because the respective images should be connected satisfying both surjection and injection, and there is no conceptual supremacy existing between these images. It is to be noted that the mappings to be constructed

here are the digital version of the bijection. In the base technology, a pixel is specified by a co-ordinate point.

The mapping of the source subimage (a subimage of a source image) to the destination subimage (a subimage of a destination image) is represented by $f^{(m,s)}: I/2^{n-m} \times I/2^{n-m} \rightarrow I/2^{n-m} \times I/2^{n-m}$ ($s = 0, 1, \dots$), where $f_{(i,j)}^{(m,s)} = (k,l)$ means that $p_{(i,j)}^{(m,s)}$ of the source image is mapped to $q_{(k,l)}^{(m,s)}$ of the destination image.

For simplicity, when $f(i,j) = (k,1)$ holds, a pixel $q_{(k,1)}$ is denoted by $q_{f(i,j)}$.

When the data sets are discrete as image pixels (grid points) treated in the base technology, the definition of bijectivity is important. Here, the bijection will be defined in the following manner, where i, j, k and l are all integers. First, a square region R defined on the source image plane is considered

$$p_{(i,j)}^{(m,s)} p_{(i+1,j)}^{(m,s)} p_{(i+1,j+1)}^{(m,s)} p_{(i,j+1)}^{(m,s)} \quad \text{--- (4)}$$

where $i = 0, \dots, 2^m-1$, and $j = 0, \dots, 2^m-1$. The edges of R are directed as follows:

$$\overrightarrow{p_{(i,j)}^{(m,s)} p_{(i+1,j)}^{(m,s)}}, \overrightarrow{p_{(i+1,j)}^{(m,s)} p_{(i+1,j+1)}^{(m,s)}}, \overrightarrow{p_{(i+1,j+1)}^{(m,s)} p_{(i,j+1)}^{(m,s)}} \text{ and } \overrightarrow{p_{(i,j+1)}^{(m,s)} p_{(i,j)}^{(m,s)}} \quad \text{--- (5)}$$

This square region R will be mapped by f to a quadrilateral on the destination image plane:

$$q_{f(i,j)}^{(m,s)} q_{f(i+1,j)}^{(m,s)} q_{f(i+1,j+1)}^{(m,s)} q_{f(i,j+1)}^{(m,s)} \quad \text{--- (6)}$$

This mapping $f^{(m,s)}(R)$, that is,

$$f^{(m,s)}(R) = f^{(m,s)}(P_{(i,j)}^{(m,s)} P_{(i+1,j)}^{(m,s)} P_{(i+1,j+1)}^{(m,s)} P_{(i,j+1)}^{(m,s)}) = q_{f(i,j)}^{(m,s)} q_{f(i+1,j)}^{(m,s)} q_{f(i+1,j+1)}^{(m,s)} q_{f(i,j+1)}^{(m,s)}$$

should satisfy the following bijectivity conditions (referred to as BC hereinafter):

1. The edges of the quadrilateral $f^{(m,s)}(R)$ should not 5 intersect one another.

2. The orientation of the edges of $f^{(m,s)}(R)$ should be the same as that of R (clockwise in the case shown in Fig. 2, described below).

3. As a relaxed condition, a retraction mapping is allowed.

10 Without a certain type of a relaxed condition as in, for example, condition 3 above, there would be no mappings which completely satisfy the BC other than a trivial identity mapping. Here, the length of a single edge of $f^{(m,s)}(R)$ may be zero. Namely, $f^{(m,s)}(R)$ may be a triangle. However, $f^{(m,s)}(R)$

15 is not allowed to be a point or a line segment having area zero. Specifically speaking, if Fig. 2R is the original quadrilateral, Figs. 2A and 2D satisfy the BC while Figs 2B, 2C and 2E do not satisfy the BC.

In actual implementation, the following condition may be

20 further imposed to easily guarantee that the mapping is surjective. Namely, each pixel on the boundary of the source image is mapped to the pixel that occupies the same location at the destination image. In other words, $f(i,j) = (i,j)$ (on the four lines of $i=0$, $i=2^m-1$, $j=0$, $j=2^m-1$). This condition

will be hereinafter referred to as an additional condition.

[1. 3. 2] Energy of mapping

[1. 3. 2. 1] Cost related to the pixel intensity

5 The energy of the mapping f is defined. An objective here is to search a mapping whose energy becomes minimum. The energy is determined mainly by the difference in the intensity between the pixel of the source image and its corresponding pixel of the destination image. Namely, the energy $C_{(i,j)}^{(m,s)}$ of
 10 the mapping $f^{(m,s)}$ at (i,j) is determined by the following equation (7).

$$C_{(i,j)}^{(m,s)} = |V(p_{(i,j)}^{(m,s)}) - V(q_{f(i,j)}^{(m,s)})|^2 \quad \text{--- (7)}$$

where $V(p_{(i,j)}^{(m,s)})$ and $V(q_{f(i,j)}^{(m,s)})$ are the intensity values of the pixels $p_{(i,j)}^{(m,s)}$ and $q_{f(i,j)}^{(m,s)}$, respectively. The total energy $C^{(m,s)}$ of
 15 f is a matching evaluation equation, and can be defined as the sum of $C_{(i,j)}^{(m,s)}$ as shown in the following equation (8).

$$C_f^{(m,s)} = \sum_{i=0}^{i=2^m-1} \sum_{j=0}^{j=2^m-1} C_{(i,j)}^{(m,s)} \quad \text{--- (8)}$$

[1. 3. 2. 2] Cost related to the locations of the pixel for

smooth mapping

In order to obtain smooth mappings, another energy D_f for the mapping is introduced. The energy D_f is determined by

the locations of $p_{(i,j)}^{(m,s)}$ and $q_{f(i,j)}^{(m,s)}$ ($i=0, 1, \dots, 2^m-1$, $j=0, 1, \dots, 2^m-1$), regardless of the intensity of the pixels. The energy $D_{(i,j)}^{(m,s)}$ of the mapping $f^{(m,s)}$ at a point (i,j) is determined by the following equation (9).

5
$$D_{(i,j)}^{(m,s)} = \eta E_{0(i,j)}^{(m,s)} + E_{1(i,j)}^{(m,s)} \quad \text{--- (9)}$$

where the coefficient parameter η which is equal to or greater than 0 is a real number. And we have

$$E_{0(i,j)}^{(m,s)} = \|(i,j) - f^{(m,s)}(i,j)\|^2 \quad \text{--- (10)}$$

$$E_{1(i,j)}^{(m,s)} = \sum_{i'=i-1}^i \sum_{j'=j-1}^j \|(f^{(m,s)}(i,j) - (i,j)) - (f^{(m,s)}(i',j') - (i',j'))\|^2 / 4 \quad \text{--- (11)}$$

10 where

$$\|(x,y)\| = \sqrt{x^2 + y^2} \quad \text{--- (12),}$$

i' and j' are integers and $f(i',j')$ is defined to be zero for $i' < 0$ and $j' < 0$. E_0 is determined by the distance between (i,j) and $f(i,j)$. E_0 prevents a pixel from being mapped to a pixel

15 too far away from it. However, as explained below, E_0 can be replaced by another energy function. E_1 ensures the smoothness of the mapping. E_1 represents a distance between the displacement of $p(i,j)$ and the displacement of its neighboring points. Based on the above consideration, another 20 evaluation equation for evaluating the matching, or the energy D_f is determined by the following equation:

$$D_f^{(m,s)} = \sum_{i=0}^{i=2^m-1} \sum_{j=0}^{j=2^m-1} D_{(i,j)}^{(m,s)} \quad \text{--- (13)}$$

[1. 3. 2. 3] Total energy of the mapping

The total energy of the mapping, that is, a combined

5 evaluation equation which relates to the combination of a plurality of evaluations, is defined as $\lambda C_f^{(m,s)} + D_f^{(m,s)}$, where $\lambda \geq 0$ is a real number. The goal is to detect a state in which the combined evaluation equation has an extreme value, namely, to find a mapping which gives the minimum energy expressed by the
10 following:

$$\min_f \{\lambda C_f^{(m,s)} + D_f^{(m,s)}\} \quad \text{--- (14)}$$

Care must be exercised in that the mapping becomes an identity mapping if $\lambda=0$ and $\eta=0$ (i.e., $f^{(m,s)}(i,j)=(i,j)$ for all $i=0,1,\dots,2^m-1$ and $j=0,1,\dots,2^m-1$). As will be described later,

15 the mapping can be gradually modified or transformed from an identity mapping since the case of $\lambda=0$ and $\eta=0$ is evaluated at the outset in the base technology. If the combined evaluation equation is defined as $C_f^{(m,s)} + \lambda D_f^{(m,s)}$ where the original position of λ is changed as such, the equation with $\lambda=0$ and $\eta=0$ will be
20 $C_f^{(m,s)}$ only. As a result thereof, pixels would randomly matched to each other only because their pixel intensities are close, thus making the mapping totally meaningless. Transforming the

mapping based on such a meaningless mapping makes no sense.

Thus, the coefficient parameter is so determined that the identity mapping is initially selected for the evaluation as the best mapping.

5 Similar to this base technology, differences in the pixel intensity and smoothness are considered in a technique called "optical flow" that is known in the art. However, the optical flow technique cannot be used for image transformation since the optical flow technique takes into account only the
10 local movement of an object. However, global correspondence can also be detected by utilizing the critical point filter according to the base technology.

[1. 3. 3] Determining the mapping with multiresolution

15 A mapping f_{\min} which gives the minimum energy and satisfies the BC is searched by using the multiresolution hierarchy. The mapping between the source subimage and the destination subimage at each level of the resolution is computed. Starting from the top of the resolution hierarchy
20 (i.e., the coarsest level), the mapping is determined at each resolution level, and where possible, mappings at other levels are considered. The number of candidate mappings at each level is restricted by using the mappings at an upper (i.e., coarser) level of the hierarchy. More specifically speaking,

in the course of determining a mapping at a certain level, the mapping obtained at the coarser level by one is imposed as a sort of constraint condition.

We thus define a parent and child relationship between 5 resolution levels. When the following equation (15) holds,

$$(i', j') = \left(\left\lfloor \frac{i}{2} \right\rfloor, \left\lfloor \frac{j}{2} \right\rfloor \right) \quad \text{--- (15),}$$

where $\lfloor x \rfloor$ denotes the largest integer not exceeding x ,

$p_{(i',j')}^{(m-1,s)}$ and $q_{(i',j')}^{(m-1,s)}$ are respectively called the parents of $p_{(i,j)}^{(m,s)}$

and $q_{(i,j)}^{(m,s)}$. Conversely, $p_{(i,j)}^{(m,s)}$ and $q_{(i,j)}^{(m,s)}$ are the child of $p_{(i',j')}^{(m-1,s)}$

10 and the child of $q_{(i',j')}^{(m-1,s)}$, respectively. A function $\text{parent}(i, j)$ is defined by the following equation (16):

$$\text{parent}(i, j) = \left(\left\lfloor \frac{i}{2} \right\rfloor, \left\lfloor \frac{j}{2} \right\rfloor \right) \quad \text{--- (16)}$$

Now, a mapping between $p_{(i,j)}^{(m,s)}$ and $q_{(k,l)}^{(m,s)}$ is determined by computing the energy and finding the minimum thereof. The

15 value of $f^{(m,s)}(i, j) = (k, l)$ is determined as follows using $f^{(m-1,s)}$ ($m=1, 2, \dots, n$). First of all, a condition is imposed that $q_{(k,l)}^{(m,s)}$ should lie inside a quadrilateral defined by the

following definitions (17) and (18). Then, the applicable mappings are narrowed down by selecting ones that are thought 20 to be reasonable or natural among them satisfying the BC.

$$q_{g^{(m,s)}(i-1,j-1)}^{(m,s)} q_{g^{(m,s)}(i-1,j+1)}^{(m,s)} q_{g^{(m,s)}(i+1,j+1)}^{(m,s)} q_{g^{(m,s)}(i+1,j-1)}^{(m,s)} \quad \text{--- (17)}$$

where

$$g^{(m,s)}(i,j) = f^{(m-1,s)}(\text{parent}(i,j)) + f^{(m-1,s)}(\text{parent}(i,j) + (1,1)) \quad \text{--- (18)}$$

The quadrilateral defined above is hereinafter referred to as the inherited quadrilateral of $p_{(i,j)}^{(m,s)}$. The pixel 5 minimizing the energy is sought and obtained inside the inherited quadrilateral.

Fig. 3 illustrates the above-described procedures. The pixels A, B, C and D of the source image are mapped to A', B', C' and D' of the destination image, respectively, at the (m-

10 1)th level in the hierarchy. The pixel $p_{(i,j)}^{(m,s)}$ should be mapped to the pixel $q_{f^{(m)}(i,j)}^{(m,s)}$ which exists inside the inherited quadrilateral A'B'C'D'. Thereby, bridging from the mapping at the (m-1)th level to the mapping at the m-th level is achieved.

15 The energy E_0 defined above may now be replaced by the following equations (19) and (20) :

$$E_{0(i,j)} = \|f^{(m,0)}(i,j) - g^{(m)}(i,j)\|^2 \quad \text{--- (19)}$$

$$E_{0(i,j)} = \|f^{(m,s)}(i,j) - f^{(m,s-1)}(i,j)\|^2, (1 \leq i) \quad \text{--- (20)}$$

20 for computing the submapping $f^{(m,0)}$ and the submapping $f^{(m,s)}$ at the m-th level, respectively.

In this manner, a mapping which maintains a low energy of all the submappings is obtained. Using the equation (20)

makes the submappings corresponding to the different critical points associated to each other within the same level in order that the subimages can have high similarity. The equation (19) represents the distance between $f^{(m,s)}(i,j)$ and the 5 location where (i,j) should be mapped when regarded as a part of a pixel at the $(m-1)$ the level.

When there is no pixel satisfying the BC inside the inherited quadrilateral $A'B'C'D'$, the following steps are taken. First, pixels whose distance from the boundary of 10 $A'B'C'D'$ is L (at first, $L=1$) are examined. If a pixel whose energy is the minimum among them satisfies the BC, then this pixel will be selected as a value of $f^{(m,s)}(i,j)$. L is increased until such a pixel is found or L reaches its upper bound $L_{\max}^{(m)}$. $L_{\max}^{(m)}$ is fixed for each level m . If no pixel is 15 found at all, the third condition of the BC is ignored temporarily and such mappings that caused the area of the transformed quadrilateral to become zero (a point or a line) will be permitted so as to determine $f^{(m,s)}(i,j)$. If such a pixel is still not found, then the first and the second 20 conditions of the BC will be removed.

Multiresolution approximation is essential to determining the global correspondence of the images while preventing the mapping from being affected by small details of the images. Without the multiresolution approximation, it is

impossible to detect a correspondence between pixels whose distances are large. In the case where the multiresolution approximation is not available, the size of an image will generally be limited to a very small size, and only tiny

5 changes in the images can be handled. Moreover, imposing smoothness on the mapping usually makes it difficult to find the correspondence of such pixels. That is because the energy of the mapping from one pixel to another pixel which is far therefrom is high. On the other hand, the multiresolution

10 approximation enables finding the approximate correspondence of such pixels. This is because the distance between the pixels is small at the upper (coarser) level of the hierarchy of the resolution.

15 [1. 4] Automatic determination of the optimal parameter values

One of the main deficiencies of the existing image matching techniques lies in the difficulty of parameter adjustment. In most cases, the parameter adjustment is performed manually and it is extremely difficult to select the

20 optimal value. However, according to the base technology, the optimal parameter values can be obtained completely automatically.

The systems according to this base technology include two parameters, namely, λ and η , where λ and η represent the

weight of the difference of the pixel intensity and the stiffness of the mapping, respectively. In order to automatically determine these parameters, they are initially set to 0. First, λ is gradually increased from $\lambda=0$ while η is fixed at 0. As λ becomes larger and the value of the combined evaluation equation (equation (14)) is minimized, the value of $C_f^{(m,s)}$ for each submapping generally becomes smaller. This

basically means that the two images are matched better.

However, if λ exceeds the optimal value, the following

10 phenomena occur:

1. Pixels which should not be corresponded are erroneously corresponded only because their intensities are close.
2. As a result, correspondence between images becomes inaccurate, and the mapping becomes invalid.

15 3. As a result, $D_f^{(m,s)}$ in equation (14) tends to increase abruptly.

4. As a result, since the value of equation (14) tends to increase abruptly, $f^{(m,s)}$ changes in order to suppress the abrupt increase of $D_f^{(m,s)}$. As a result, $C_f^{(m,s)}$ increases.

20 Therefore, a threshold value at which $C_f^{(m,s)}$ turns to an increase from a decrease is detected while a state in which equation (14) takes the minimum value with λ being increased is kept. Such λ is determined as the optimal value at $\eta=0$.

Next, the behavior of $C_f^{(m,s)}$ is examined while η is increased gradually, and η will be automatically determined by a method described later. λ will then again be determined corresponding to such an automatically determined η .

5 The above-described method resembles the focusing mechanism of human visual systems. In the human visual systems, the images of the respective right eye and left eye are matched while moving one eye. When the objects are clearly recognized, the moving eye is fixed.

10

[1. 4. 1] Dynamic determination of λ

Initially, λ is increased from 0 at a certain interval, and a subimage is evaluated each time the value of λ changes.

As shown in equation (14), the total energy is defined by

15 $\lambda C_f^{(m,s)} + D_f^{(m,s)}$. $D_f^{(m,s)}$ in equation (9) represents the smoothness and theoretically becomes minimum when it is the identity mapping. E_0 and E_1 increase as the mapping is further distorted. Since E_1 is an integer, 1 is the smallest step of $D_f^{(m,s)}$. Thus, it is impossible to change the mapping to reduce 20 the total energy unless a changed amount (reduction amount) of the current $\lambda C_{(i,j)}^{(m,s)}$ is equal to or greater than 1. Since $D_f^{(m,s)}$ increases by more than 1 accompanied by the change of the mapping, the total energy is not reduced unless $\lambda C_{(i,j)}^{(m,s)}$ is

reduced by more than 1.

Under this condition, it is shown that $C_{(i,j)}^{(m,s)}$ decreases in normal cases as λ increases. The histogram of $C_{(i,j)}^{(m,s)}$ is denoted as $h(l)$, where $h(l)$ is the number of pixels whose energy $C_{(i,j)}^{(m,s)}$ is l^2 . In order that $\lambda l^2 \geq 1$, for example, the case of $l^2=1/\lambda$ is considered. When λ varies from λ_1 to λ_2 , a number of pixels (denoted A) expressed by the following equation (21):

$$A = \sum_{l=\left[\frac{1}{\lambda_2}\right]}^{\left[\frac{1}{\lambda_1}\right]} h(l) \cong \int_{l=\frac{1}{\lambda_2}}^{\frac{1}{\lambda_1}} h(l) dl = - \int_{\lambda_2}^{\lambda_1} h(l) \frac{1}{\lambda^{3/2}} d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{h(l)}{\lambda^{3/2}} d\lambda \quad --- (21)$$

changes to a more stable state having the energy shown in equation (22):

$$C_f^{(m,s)} - l^2 = C_f^{(m,s)} - \frac{1}{\lambda} \quad --- (22)$$

Here, it is assumed that the energy of these pixels is approximated to be zero. This means that the value of $C_{(i,j)}^{(m,s)}$ changes by:

$$15 \quad \partial C_f^{(m,s)} = -\frac{A}{\lambda} \quad --- (23)$$

As a result, equation (24) holds.

$$\frac{\partial C_f^{(m,s)}}{\partial \lambda} = -\frac{h(l)}{\lambda^{5/2}} \quad --- (24)$$

Since $h(l) > 0$, $C_f^{(m,s)}$ decreases in the normal case. However, when λ exceeds the optimal value, the above phenomenon, that

is, an increase in $C_f^{(m,s)}$ occurs. The optimal value of λ is determined by detecting this phenomenon.

When

$$h(l) = Hl^k = \frac{H}{\lambda^{k/2}} \quad \text{--- (25)}$$

5 is assumed, where both $H (H>0)$ and k are constants, the equation (26) holds:

$$\frac{\partial C_f^{(m,s)}}{\partial \lambda} = -\frac{H}{\lambda^{5/2+k/2}} \quad \text{--- (26)}$$

Then, if $k \neq -3$, the following equation (27) holds:

$$C_f^{(m,s)} = C + \frac{H}{(3/2+k/2)\lambda^{3/2+k/2}} \quad \text{--- (27)}$$

10 The equation (27) is a general equation of $C_f^{(m,s)}$ (where C is a constant).

When detecting the optimal value of λ , the number of pixels violating the BC may be examined for safety. In the course of determining a mapping for each pixel, the 15 probability of violating the BC is assumed as a value p_0 here. In this case, since

$$\frac{\partial A}{\partial \lambda} = \frac{h(l)}{\lambda^{3/2}} \quad \text{--- (28)}$$

holds, the number of pixels violating the BC increases at a rate of:

$$20 \quad B_0 = \frac{h(l)p_0}{\lambda^{3/2}} \quad \text{--- (29)}$$

Thus,

$$\frac{B_0 \lambda^{3/2}}{p_0 h(l)} = 1 \quad \text{--- (30)}$$

is a constant. If it is assumed that $h(l) = Hl^k$, the following equation (31), for example,

5 $B_0 \lambda^{3/2+k/2} = p_0 H \quad \text{--- (31)}$

becomes a constant. However, when λ exceeds the optimal value, the above value of equation (31) increases abruptly.

By detecting this phenomenon, i.e. whether or not the value of $B_0 \lambda^{3/2+k/2} / 2^m$ exceeds an abnormal value $B_{0\text{thres}}$, the optimal value

10 of λ can be determined. Similarly, whether or not the value of $B_1 \lambda^{3/2+k/2} / 2^m$ exceeds an abnormal value $B_{1\text{thres}}$ can be used to check for an increasing rate B_1 of pixels violating the third condition of the BC. The reason why the factor 2^m is introduced here will be described at a later stage. This

15 system is not sensitive to the two threshold values $B_{0\text{thres}}$ and $B_{1\text{thres}}$. The two threshold values $B_{0\text{thres}}$ and $B_{1\text{thres}}$ can be used to detect excessive distortion of the mapping which may not be detected through observation of the energy $C_f^{(m,s)}$.

In the experimentation, when λ exceeded 0.1 the 20 computation of $f^{(m,s)}$ was stopped and the computation of $f^{(m,s+1)}$ was started. That is because the computation of submappings is affected by a difference of only 3 out of 255 levels in

pixel intensity when $\lambda > 0.1$ and it is then difficult to obtain a correct result.

[1. 4. 2] Histogram $h(l)$

5 The examination of $C_f^{(m,s)}$ does not depend on the histogram $h(l)$, however, the examination of the BC and its third condition may be affected by $h(l)$. When $(\lambda, C_f^{(m,s)})$ is actually plotted, k is usually close to 1. In the experiment, $k=1$ is used, that is, $B_0\lambda^2$ and $B_1\lambda^2$ are examined. If the true value 10 of k is less than 1, $B_0\lambda^2$ and $B_1\lambda^2$ are not constants and increase gradually by a factor of $\lambda^{(1-k)/2}$. If $h(l)$ is a constant, the factor is, for example, $\lambda^{1/2}$. However, such a difference can be absorbed by setting the threshold $B_{0\text{thres}}$ appropriately.

15 Let us model the source image by a circular object, with its center at (x_0, y_0) and its radius r , given by:

$$p(i, j) = \begin{cases} \frac{255}{r} c(\sqrt{(i-x_0)^2 + (j-y_0)^2}) \dots (\sqrt{(i-x_0)^2 + (j-y_0)^2} \leq r) \\ 0 \dots (\text{otherwise}) \end{cases}$$

--- (32)

and the destination image given by:

$$20 q(i, j) = \begin{cases} \frac{255}{r} c(\sqrt{(i-x_1)^2 + (j-y_1)^2}) \dots (\sqrt{(i-x_1)^2 + (j-y_1)^2} \leq r) \\ 0 \dots (\text{otherwise}) \end{cases}$$

--- (33)

with its center at (x_1, y_1) and radius r . In the above, let $c(x)$ have the form of $c(x)=x^k$. When the centers (x_0, y_0) and (x_1, y_1) are sufficiently far from each other, the histogram

5 $h(l)$ is then in the form:

$$h(l) \propto rl^k (k \neq 0) \quad \text{--- (34)}$$

When $k=1$, the images represent objects with clear boundaries embedded in the background. These objects become darker toward their centers and brighter toward their

10 boundaries. When $k=-1$, the images represent objects with vague boundaries. These objects are brightest at their centers, and become darker toward their boundaries. Without much loss of generality, it suffices to state that objects in images are generally between these two types of objects.

15 Thus, choosing k such that $-1 \leq k \leq 1$ can cover most cases and the equation (27) is generally a decreasing function for this range.

As can be observed from the above equation (34), attention must be directed to the fact that r is influenced by

20 the resolution of the image, that is, r is proportional to 2^m . This is the reason for the factor 2^m being introduced in the above section [1.4.1].

[1. 4. 3] Dynamic determination of η

The parameter η can also be automatically determined in a similar manner. Initially, η is set to zero, and the final mapping $f^{(n)}$ and the energy $C_f^{(n)}$ at the finest resolution are computed. Then, after η is increased by a certain value $\Delta\eta$,

- 5 the final mapping $f^{(n)}$ and the energy $C_f^{(n)}$ at the finest resolution are again computed. This process is repeated until the optimal value of η is obtained. η represents the stiffness of the mapping because it is a weight of the following equation (35):

10
$$E_{0(i,j)}^{(m,s)} = \|f^{(m,s)}(i,j) - f^{(m,s-1)}(i,j)\|^2 \quad \text{--- (35)}$$

If η is zero, $D_f^{(n)}$ is determined irrespective of the previous submapping, and the present submapping may be elastically deformed and become too distorted. On the other hand, if η is a very large value, $D_f^{(n)}$ is almost completely determined by the immediately previous submapping. The submappings are then very stiff, and the pixels are mapped to almost the same locations. The resulting mapping is therefore the identity mapping. When the value of η increases from 0, $C_f^{(n)}$ gradually decreases as will be described later. However, 20 when the value of η exceeds the optimal value, the energy starts increasing as shown in Fig. 4. In Fig. 4, the x-axis represents η , and y-axis represents C_f .

The optimum value of η which minimizes $C_f^{(n)}$ can be obtained in this manner. However, since various elements affect this computation as compared to the case of λ , $C_f^{(n)}$ changes while slightly fluctuating. This difference is caused 5 because a submapping is re-computed once in the case of λ whenever an input changes slightly, whereas all the submappings must be re-computed in the case of η . Thus, whether the obtained value of $C_f^{(n)}$ is the minimum or not cannot be determined as easily. When candidates for the minimum 10 value are found, the true minimum needs to be searched by setting up further finer intervals.

[1. 5] Supersampling

When deciding the correspondence between the pixels, the 15 range of $f^{(m,s)}$ can be expanded to $R \times R$ (R being the set of real numbers) in order to increase the degree of freedom. In this case, the intensity of the pixels of the destination image is interpolated, to provide $f^{(m,s)}$ having an intensity at non-integer points:

20 $V(q_{f^{(m,s)}(i,j)}^{(m,s)})$ --- (36)

That is, supersampling is performed. In an example implementation, $f^{(m,s)}$ may take integer and half integer values, and

$$V(q_{(i,j)+(0.5,0.5)}^{(m,s)}) \quad --- \quad (37)$$

is given by

$$(V(q_{(i,j)}^{(m,s)}) + V(q_{(i,j)+(1,1)}^{(m,s)})) / 2 \quad --- \quad (38)$$

5 [1. 6] Normalization of the pixel intensity of each image

When the source and destination images contain quite different objects, the raw pixel intensity may not be used to compute the mapping because a large difference in the pixel intensity causes excessively large energy $C_f^{(m,s)}$ and thus making 10 it difficult to obtain an accurate evaluation.

For example, a matching between a human face and a cat's face is computed as shown in Figs. 20(a) and 20(b). The cat's face is covered with hair and is a mixture of very bright pixels and very dark pixels. In this case, in order to 15 compute the submappings of the two faces, subimages are normalized. That is, the darkest pixel intensity is set to 0 while the brightest pixel intensity is set to 255, and other pixel intensity values are obtained using linear interpolation.

20

[1. 7] Implementation

In an example implementation, a heuristic method is utilized wherein the computation proceeds linearly as the

source image is scanned. First, the value of $f^{(m,s)}$ is determined at the top leftmost pixel $(i,j)=(0,0)$. The value of each $f^{(m,s)}(i,j)$ is then determined while i is increased by one at each step. When i reaches the width of the image, j is

5 increased by one and i is reset to zero. Thereafter,

$f^{(m,s)}(i,j)$ is determined while scanning the source image.

Once pixel correspondence is determined for all the points, it means that a single mapping $f^{(m,s)}$ is determined.

When a corresponding point $q_{f(i,j)}$ is determined for $p_{(i,j)}$,

10 a corresponding point $q_{f(i,j+1)}$ of $p_{(i,j+1)}$ is determined next.

The position of $q_{f(i,j+1)}$ is constrained by the position of $q_{f(i,j)}$ since the position of $q_{f(i,j+1)}$ satisfies the BC. Thus, in this system, a point whose corresponding point is determined earlier is given higher priority. If the situation continues

15 in which $(0,0)$ is always given the highest priority, the final mapping might be unnecessarily biased. In order to avoid this bias, $f^{(m,s)}$ is determined in the following manner in the base technology.

First, when $(s \bmod 4)$ is 0, $f^{(m,s)}$ is determined starting

20 from $(0,0)$ while gradually increasing both i and j . When $(s \bmod 4)$ is 1, $f^{(m,s)}$ is determined starting from the top

rightmost location while decreasing i and increasing j . When

$(s \bmod 4)$ is 2, $f^{(m,s)}$ is determined starting from the bottom

rightmost location while decreasing both i and j . When $(s \bmod$

4) is 3, $f^{(m,s)}$ is determined starting from the bottom leftmost location while increasing i and decreasing j . Since a concept such as the submapping, that is, a parameter s , does not exist in the finest n -th level, $f^{(m,s)}$ is computed continuously in two 5 directions on the assumption that $s=0$ and $s=2$.

In this implementation, the values of $f^{(m,s)}(i,j)$ ($m=0, \dots, n$) that satisfy the BC are chosen as much as possible from the candidates (k,l) by imposing a penalty on the candidates violating the BC. The energy $D_{(k,l)}$ of a candidate 10 that violates the third condition of the BC is multiplied by φ and that of a candidate that violates the first or second condition of the BC is multiplied by ψ . In this implementation, $\varphi=2$ and $\psi=100000$ are used.

In order to check the above-mentioned BC, the following 15 test may be performed as the procedure when determining $(k,l)=f^{(m,s)}(i,j)$. Namely, for each grid point (k,l) in the inherited quadrilateral of $f^{(m,s)}(i,j)$, whether or not the z -component of the outer product of

$$W = \vec{A} \times \vec{B} \quad \text{--- (39)}$$

20 is equal to or greater than 0 is examined, where

$$\vec{A} = \overrightarrow{q_{f^{(m,s)}(i,j-1)}^{(m,s)} q_{f^{(m,s)}(i+1,j-1)}^{(m,s)}} \quad \text{--- (40)}$$

$$\vec{B} = \overrightarrow{q_{f^{(m,s)}(i,j-1)}^{(m,s)} q_{(k,l)}^{(m,s)}} \quad \text{--- (41)}$$

Here, the vectors are regarded as 3D vectors and the z -axis is

defined in the orthogonal right-hand coordinate system. When W is negative, the candidate is imposed with a penalty by multiplying $D_{(k,l)}^{(m,s)}$ by ψ so that it is not as likely to be selected.

5 Figs. 5(a) and 5(b) illustrate the reason why this condition is inspected. Fig. 5(a) shows a candidate without a penalty and Fig. 5(b) shows one with a penalty. When determining the mapping $f^{(m,s)}(i, j+1)$ for the adjacent pixel at $(i, j+1)$, there is no pixel on the source image plane that 10 satisfies the BC if the z -component of W is negative because then $q_{(k,l)}^{(m,s)}$ passes the boundary of the adjacent quadrilateral.

[1. 7. 1] The order of submappings

In this implementation, $\sigma(0)=0$, $\sigma(1)=1$, $\sigma(2)=2$, $\sigma(3)=3$, 15 $\sigma(4)=0$ are used when the resolution level is even, while $\sigma(0)=3$, $\sigma(1)=2$, $\sigma(2)=1$, $\sigma(3)=0$, $\sigma(4)=3$ are used when the resolution level is odd. Thus, the submappings are shuffled to some extent. It is to be noted that the submappings are primarily of four types, and s may be any of 0 to 3. However, 20 a processing with $s=4$ is used in this implementation for a reason to be described later.

[1. 8] Interpolations

After the mapping between the source and destination

MN-70022

images is determined, the intensity values of the corresponding pixels are interpolated. In the implementation, trilinear interpolation is used. Suppose that a square

$P(i,j)P(i+1,j)P(i+1,j+1)P(i,j+1)$ on the source image plane is mapped to

- 5 a quadrilateral $q_f(i,j)q_f(i+1,j)q_f(i+1,j+1)q_f(i,j+1)$ on the destination image plane. For simplicity, the distance between the image planes is assumed to be 1. The intermediate image pixels $r(x,y,t)$ ($0 \leq x \leq N-1, 0 \leq y \leq M-1$) whose distance from the source image plane is t ($0 \leq t \leq 1$) are obtained as follows. First, the
 10 location of the pixel $r(x,y,t)$, where $x,y,t \in R$, is determined by equation (42):

$$(x,y) = (1-dx)(1-dy)(1-t)(i,j) + (1-dx)(1-dy)tf(i,j) \\ + dx(1-dy)(1-t)(i+1,j) + dx(1-dy)tf(i+1,j) \quad --- \quad (42) \\ + (1-dx)dy(1-t)(i,j+1) + (1-dx)dytf(i,j+1) \\ + dxdy(1-t)(i+1,j+1) + dxdytf(i+1,j+1)$$

The value of the pixel intensity at $r(x,y,t)$ is then determined by equation (43):

$$15 \quad V(r(x,y,t)) = (1-dx)(1-dy)(1-t)V(p_{(i,j)}) + (1-dx)(1-dy)tV(q_{f(i,j)}) \\ + dx(1-dy)(1-t)V(p_{(i+1,j)}) + dx(1-dy)tV(q_{f(i+1,j)}) \\ + (1-dx)dy(1-t)V(p_{(i,j+1)}) + (1-dx)dytV(q_{f(i,j+1)}) \\ + dxdy(1-t)V(p_{(i+1,j+1)}) + dxdytV(q_{f(i+1,j+1)}) \quad --- \quad (43)$$

where dx and dy are parameters varying from 0 to 1.

[1. 9] Mapping to which constraints are imposed

- 20 So far, the determination of a mapping in which no

constraints are imposed has been described. However, if a correspondence between particular pixels of the source and destination images is provided in a predetermined manner, the mapping can be determined using such correspondence as a 5 constraint.

The basic idea is that the source image is roughly deformed by an approximate mapping which maps the specified pixels of the source image to the specified pixels of the destination image and thereafter a mapping f is accurately 10 computed.

First, the specified pixels of the source image are mapped to the specified pixels of the destination image, then the approximate mapping that maps other pixels of the source image to appropriate locations are determined. In other words, 15 the mapping is such that pixels in the vicinity of a specified pixel are mapped to locations near the position to which the specified one is mapped. Here, the approximate mapping at the m -th level in the resolution hierarchy is denoted by $F^{(m)}$.

The approximate mapping F is determined in the following 20 manner. First, the mappings for several pixels are specified. When n_s pixels

$$p(i_0, j_0), p(i_1, j_1), \dots, p(i_{n_s-1}, j_{n_s-1}) \quad \text{--- (44)}$$

of the source image are specified, the following values in the equation (45) are determined.

$$\begin{aligned}
 F^{(n)}(i_0, j_0) &= (k_0, l_0), \\
 F^{(n)}(i_1, j_1) &= (k_1, l_1), \dots, \\
 F^{(n)}(i_{n_s-1}, j_{n_s-1}) &= (k_{n_s-1}, l_{n_s-1})
 \end{aligned}
 \quad \text{--- (45)}$$

For the remaining pixels of the source image, the amount of displacement is the weighted average of the displacement of $p(i_h, j_h)$ ($h=0, \dots, n_s-1$). Namely, a pixel $p_{(i,j)}$ is mapped to the 5 following pixel (expressed by the equation (46)) of the destination image.

$$F^{(m)}(i, j) = \frac{(i, j) + \sum_{h=0}^{n_s-1} (k_h - i_h, l_h - j_h) weight_h(i, j)}{2^{n-m}} \quad \text{--- (46)}$$

where

$$weight_h(i, j) = \frac{1/\|(i_h - i, j_h - j)\|^2}{total_weight(i, j)} \quad \text{--- (47)}$$

10 where

$$total_weight(i, j) = \sum_{h=0}^{n_s-1} 1/\|(i_h - i, j_h - j)\|^2 \quad \text{--- (48)}$$

Second, the energy $D_{(i,j)}^{(m,s)}$ of the candidate mapping f is changed so that a mapping f similar to $F^{(m)}$ has a lower energy.

Precisely speaking, $D_{(i,j)}^{(m,s)}$ is expressed by the equation (49):

$$D_{(i,j)}^{(m,s)} = E_{0(i,j)}^{(m,s)} + \eta E_{1(i,j)}^{(m,s)} + \kappa E_{2(i,j)}^{(m,s)} \quad \text{--- (49)}$$

where

$$E_{2(i,j)}^{(m,s)} = \begin{cases} 0, & \text{if } \|F^{(m)}(i, j) - f^{(m,s)}(i, j)\|^2 \leq \left\lfloor \frac{\rho^2}{2^{2(n-m)}} \right\rfloor \\ \|F^{(m)}(i, j) - f^{(m,s)}(i, j)\|^2, & \text{otherwise} \end{cases} \quad \text{--- (50)}$$

where $\kappa, \rho \geq 0$. Finally, the resulting mapping f is determined by the above-described automatic computing process.

Note that $E_{2(i,j)}^{(m,s)}$ becomes 0 if $f^{(m,s)}(i,j)$ is sufficiently close to $F^{(m)}(i,j)$ i.e., the distance therebetween is equal to 5 or less than

$$\left\lfloor \frac{\rho^2}{2^{2(n-m)}} \right\rfloor \quad \text{--- (51)}$$

This has been defined in this way because it is desirable to determine each value $f^{(m,s)}(i,j)$ automatically to fit in an appropriate place in the destination image as long as each 10 value $f^{(m,s)}(i,j)$ is close to $F^{(m)}(i,j)$. For this reason, there is no need to specify the precise correspondence in detail to have the source image automatically mapped so that the source image matches the destination image.

15 [2] Concrete Processing Procedure

The flow of a process utilizing the respective elemental techniques described in [1] will now be described.

Fig. 6 is a flowchart of the overall procedure of the base technology. Referring to Fig. 6, a source image and 20 destination image are first processed using a multiresolutional critical point filter (S1). The source image and the destination image are then matched (S2). As will be understood, the matching (S2) is not required in every case,

and other processing such as image recognition may be performed instead, based on the characteristics of the source image obtained at S1.

Fig. 7 is a flowchart showing details of the process S1 shown in Fig. 6. This process is performed on the assumption that a source image and a destination image are matched at S2. Thus, a source image is first hierarchized using a critical point filter (S10) so as to obtain a series of source hierarchical images. Then, a destination image is hierarchized in the similar manner (S11) so as to obtain a series of destination hierarchical images. The order of S10 and S11 in the flow is arbitrary, and the source image and the destination image can be generated in parallel. It may also be possible to process a number of source and destination images as required by subsequent processes.

Fig. 8 is a flowchart showing details of the process at S10 shown in Fig. 7. Suppose that the size of the original source image is $2^n \times 2^n$. Since source hierarchical images are sequentially generated from an image with a finer resolution to one with a coarser resolution, the parameter m which indicates the level of resolution to be processed is set to n (S100). Then, critical points are detected from the images $p^{(m,0)}$, $p^{(m,1)}$, $p^{(m,2)}$ and $p^{(m,3)}$ of the m -th level of resolution, using a critical point filter (S101), so that the images $p^{(m-1,0)}$,

$p^{(m-1,0)}$, $p^{(m-1,1)}$, $p^{(m-1,2)}$ and $p^{(m-1,3)}$ of the $(m-1)$ th level are generated (S102). Since $m=n$ here, $p^{(m,0)} = p^{(m,1)} = p^{(m,2)} = p^{(m,3)} = p^{(n)}$ holds and four types of subimages are thus generated from a single source image.

5 Fig. 9 shows correspondence between partial images of the m -th and those of $(m-1)$ th levels of resolution. Referring to Fig. 9, respective numeric values shown in the figure represent the intensity of respective pixels. $p^{(m,s)}$ symbolizes any one of four images $p^{(m,0)}$ through $p^{(m,3)}$, and when 10 generating $p^{(m-1,0)}$, $p^{(m,0)}$ is used from $p^{(m,s)}$. For example, as for the block shown in Fig. 9, comprising four pixels with their pixel intensity values indicated inside, images $p^{(m-1,0)}$, $p^{(m-1,1)}$, $p^{(m-1,2)}$ and $p^{(m-1,3)}$ acquire "3", "8", "6" and "10", respectively, according to the rules described in [1.2]. This 15 block at the m -th level is replaced at the $(m-1)$ th level by respective single pixels thus acquired. Therefore, the size of the subimages at the $(m-1)$ th level is $2^{m-1} \times 2^{m-1}$.

After m is decremented (S103 in Fig. 8), it is ensured that m is not negative (S104). Thereafter, the process 20 returns to S101, so that subimages of the next level of resolution, i.e., a next coarser level, are generated. The above process is repeated until subimages at $m=0$ (0-th level) are generated to complete the process at S10. The size of the subimages at the 0-th level is 1×1 .

Fig. 10 shows source hierarchical images generated at S10 in the case of $n=3$. The initial source image is the only image common to the four series followed. The four types of subimages are generated independently, depending on the type 5 of critical point. Note that the process in Fig. 8 is common to S11 shown in Fig. 7, and that destination hierarchical images are generated through a similar procedure. Then, the process at S1 in Fig. 6 is completed.

In this base technology, in order to proceed to S2 shown 10 in Fig. 6 a matching evaluation is prepared. Fig. 11 shows the preparation procedure. Referring to Fig. 11, a plurality of evaluation equations are set (S30). The evaluation equations may include the energy $C_f^{(m,s)}$ concerning a pixel 15 value, introduced in [1.3.2.1], and the energy $D_f^{(m,s)}$ concerning the smoothness of the mapping introduced in [1.3.2.2]. Next, by combining these evaluation equations, a combined evaluation equation is set (S31). Such a combined evaluation equation may be $\lambda C_{(i,j)}^{(m,s)} + D_f^{(m,s)}$. Using η introduced in [1.3.2.2], we have

$$\sum \sum (\lambda C_{(i,j)}^{(m,s)} + \eta E_{0(i,j)}^{(m,s)} + E_{1(i,j)}^{(m,s)}) \quad \text{--- (52)}$$

20 In the equation (52) the sum is taken for each i and j where i and j run through $0, 1, \dots, 2^m - 1$. Now, the preparation for matching evaluation is completed.

Fig. 12 is a flowchart showing the details of the

process of S2 shown in Fig. 6. As described in [1], the source hierarchical images and destination hierarchical images are matched between images having the same level of resolution. In order to detect global correspondence correctly, a matching is calculated in sequence from a coarse level to a fine level of resolution. Since the source and destination hierarchical images are generated using the critical point filter, the location and intensity of critical points are stored clearly even at a coarse level. Thus, the result of the global matching is superior to conventional methods.

Referring to Fig. 12, a coefficient parameter η and a level parameter m are set to 0 (S20). Then, a matching is computed between the four subimages at the m -th level of the source hierarchical images and those of the destination hierarchical images at the m -th level, so that four types of submappings $f^{(m,s)}$ ($s=0, 1, 2, 3$) which satisfy the BC and minimize the energy are obtained (S21). The BC is checked by using the inherited quadrilateral described in [1.3.3]. In that case, the submappings at the m -th level are constrained by those at the $(m-1)$ -th level, as indicated by the equations (17) and (18). Thus, the matching computed at a coarser level of resolution is used in subsequent calculation of a matching. This is called a vertical reference between different levels.

If $m=0$, there is no coarser level and this exceptional case will be described using Fig. 13.

A horizontal reference within the same level is also performed. As indicated by the equation (20) in [1.3.3], 5 $f^{(m,3)}$, $f^{(m,2)}$ and $f^{(m,1)}$ are respectively determined so as to be analogous to $f^{(m,2)}$, $f^{(m,1)}$ and $f^{(m,0)}$. This is because a situation in which the submappings are totally different seems unnatural even though the type of critical points differs so long as the critical points are originally included in the 10 same source and destination images. As can be seen from the equation (20), the closer the submappings are to each other, the smaller the energy becomes, so that the matching is then considered more satisfactory.

As for $f^{(m,0)}$, which is to be initially determined, a 15 coarser level by one may be referred to since there is no other submapping at the same level to be referred to as shown in the equation (19). In this base technology, however, a procedure is adopted such that after the submappings were obtained up to $f^{(m,3)}$, $f^{(m,0)}$ is recalculated once utilizing the 20 thus obtained submappings as a constraint. This procedure is equivalent to a process in which $s=4$ is substituted into the equation (20) and $f^{(m,4)}$ is set to $f^{(m,0)}$ anew. The above process is employed to avoid the tendency in which the degree of association between $f^{(m,0)}$ and $f^{(m,3)}$ becomes too low. This

scheme actually produced a preferable result. In addition to this scheme, the submappings are shuffled in the experiment as described in [1.7.1], so as to closely maintain the degrees of association among submappings which are originally determined 5 independently for each type of critical point. Furthermore, in order to prevent the tendency of being dependent on the starting point in the process, the location thereof is changed according to the value of s as described in [1.7].

Fig. 13 illustrates how the submapping is determined at 10 the 0-th level. Since at the 0-th level each sub-image is constituted by a single pixel, the four submappings $f^{(0,s)}$ are automatically chosen as the identity mapping. Fig. 14 shows 15 how the submappings are determined at the first level. At the first level, each of the sub-images is constituted of four pixels, which are indicated by solid lines. When a corresponding point (pixel) of the point (pixel) x in $p^{(1,s)}$ is searched within $q^{(1,s)}$, the following procedure is adopted:

1. An upper left point a , an upper right point b , a lower left point c and a lower right point d with respect to the point x 20 are obtained at the first level of resolution.
2. Pixels to which the points a to d belong at a coarser level by one, i.e., the 0-th level, are searched. In Fig. 14, the points a to d belong to the pixels A to D , respectively.

However, the pixels A to C are virtual pixels which do not

exist in reality.

3. The corresponding points A' to D' of the pixels A to D , which have already been defined at the 0-th level, are plotted in $q^{(1,s)}$. The pixels A' to C' are virtual pixels and regarded

5 to be located at the same positions as the pixels A to C .

4. The corresponding point a' to the point a in the pixel A is regarded as being located inside the pixel A' , and the point a' is plotted. Then, it is assumed that the position occupied by the point a in the pixel A (in this case, positioned at the lower right) is the same as the position occupied by the point a' in the pixel A' .

5. The corresponding points b' to d' are plotted by using the same method as the above 4 so as to produce an inherited quadrilateral defined by the points a' to d' .

15 6. The corresponding point x' of the point x is searched such that the energy becomes minimum in the inherited quadrilateral. Candidate corresponding points x' may be limited to the pixels, for instance, whose centers are included in the inherited quadrilateral. In the case shown in 20 Fig. 14, the four pixels all become candidates.

The above described is a procedure for determining the corresponding point of a given point x . The same processing is performed on all other points so as to determine the submappings. As the inherited quadrilateral is expected to

become deformed at the upper levels (higher than the second level), the pixels A' to D' will be positioned apart from one another as shown in Fig. 3.

Once the four submappings at the m -th level are

- 5 determined in this manner, m is incremented (S22 in Fig. 12). Then, when it is confirmed that m does not exceed n (S23), return to S21. Thereafter, every time the process returns to S21, submappings at a finer level of resolution are obtained until the process finally returns to S21 at which time the 10 mapping $f^{(n)}$ at the n -th level is determined. This mapping is denoted as $f^{(n)}(\eta=0)$ because it has been determined relative to $\eta=0$.

Next, to obtain the mapping with respect to other different η , η is shifted by $\Delta\eta$ and m is reset to zero (S24).

- 15 After confirming that new η does not exceed a predetermined search-stop value η_{\max} (S25), the process returns to S21 and the mapping $f^{(n)}(\eta=\Delta\eta)$ relative to the new η is obtained. This process is repeated while obtaining $f^{(n)}(\eta=i\Delta\eta)$ ($i=0,1,\dots$) at S21. When η exceeds η_{\max} , the process proceeds to S26 and the optimal 20 $\eta=\eta_{\text{opt}}$ is determined using a method described later, so as to let $f^{(n)}(\eta=\eta_{\text{opt}})$ be the final mapping $f^{(n)}$.

Fig. 15 is a flowchart showing the details of the process of S21 shown in Fig. 12. According to this flowchart,

the submappings at the m -th level are determined for a certain predetermined η . In this base technology, when determining the mappings, the optimal λ is defined independently for each submapping.

- 5 Referring to Fig. 15, s and λ are first reset to zero (S210). Then, obtained is the submapping $f^{(m,s)}$ that minimizes the energy with respect to the then λ (and, implicitly, η) (S211), and the thus obtained submapping is denoted as $f^{(m,s)}(\lambda=0)$. In order to obtain the mapping with respect to other different λ , λ is shifted by $\Delta\lambda$. After confirming that the new λ does not exceed a predetermined search-stop value λ_{\max} (S213), the process returns to S211 and the mapping $f^{(m,s)}$ ($\lambda=\Delta\lambda$) relative to the new λ is obtained. This process is repeated while obtaining $f^{(m,s)}(\lambda=i\Delta\lambda)$ ($i=0,1,\dots$). When λ exceeds λ_{\max} , the process proceeds to S214 and the optimal $\lambda=\lambda_{\text{opt}}$ is determined, so as to let $f^{(n)}(\lambda=\lambda_{\text{opt}})$ be the final mapping $f^{(m,s)}$ (S214).

Next, in order to obtain other submappings at the same level, λ is reset to zero and s is incremented (S215). After confirming that s does not exceed 4 (S216), return to S211. When $s=4$, $f^{(m,0)}$ is renewed utilizing $f^{(m,3)}$ as described above and a submapping at that level is determined.

Fig. 16 shows the behavior of the energy $C_f^{(m,s)}$

corresponding to $f^{(m,s)}(\lambda = i\Delta\lambda)$ ($i = 0, 1, \dots$) for a certain m and s

while varying λ . As described in [1.4], as λ increases, $C_f^{(m,s)}$

normally decreases but changes to increase after λ exceeds the

optimal value. In this base technology, λ in which $C_f^{(m,s)}$

5 becomes the minima is defined as λ_{opt} . As observed in Fig. 16, even if $C_f^{(m,s)}$ begins to decrease again in the range $\lambda > \lambda_{opt}$, the mapping will not be as good. For this reason, it suffices to pay attention to the first occurring minima value. In this base technology, λ_{opt} is independently determined for each

10 submapping including $f^{(n)}$.

Fig. 17 shows the behavior of the energy $C_f^{(n)}$

corresponding to $f^{(n)}(\eta = i\Delta\eta)$ ($i = 0, 1, \dots$) while varying η . Here

too, $C_f^{(n)}$ normally decreases as η increases, but $C_f^{(n)}$ changes to increase after η exceeds the optimal value. Thus, η in which

15 $C_f^{(n)}$ becomes the minima is defined as η_{opt} . Fig. 17 can be considered as an enlarged graph around zero along the horizontal axis shown in Fig. 4. Once η_{opt} is determined, $f^{(n)}$ can be finally determined.

As described above, this base technology provides

20 various merits. First, since there is no need to detect edges, problems in connection with the conventional techniques of the edge detection type are solved. Furthermore, prior

knowledge about objects included in an image is not necessitated, thus automatic detection of corresponding points is achieved. Using the critical point filter, it is possible to preserve intensity and locations of critical points even at 5 a coarse level of resolution, thus being extremely advantageous when applied to object recognition, characteristic extraction, and image matching. As a result, it is possible to construct an image processing system which significantly reduces manual labor.

10 Some further extensions to or modifications of the above-described base technology may be made as follows:

(1) Parameters are automatically determined when the matching is computed between the source and destination hierarchical images in the base technology. This method can be applied not 15 only to the calculation of the matching between the hierarchical images but also to computing the matching between two images in general.

For instance, an energy E_0 relative to a difference in the intensity of pixels and an energy E_1 relative to a 20 positional displacement of pixels between two images may be used as evaluation equations, and a linear sum of these equations, i.e., $E_{tot}=\alpha E_0+E_1$, may be used as a combined evaluation equation. While paying attention to the neighborhood of the extrema in this combined evaluation

equation, α is automatically determined. Namely, mappings which minimize E_{tot} are obtained for various α 's. Among such mappings, α at which E_{tot} takes the minimum value is defined as an optimal parameter. The mapping corresponding to this 5 parameter is finally regarded as the optimal mapping between the two images.

Many other methods are available in the course of setting up evaluation equations. For instance, a term which becomes larger as the evaluation result becomes more 10 favorable, such as $1/E_1$ and $1/E_2$, may be employed. A combined evaluation equation is not necessarily a linear sum, but an n-powered sum ($n=2, 1/2, -1, -2$, etc.), a polynomial or an arbitrary function may be employed when appropriate.

The system may employ a single parameter such as the 15 above α , two parameters such as η and λ as in the base technology, or more than two parameters. When there are more than three parameters used, they may be determined while changing one at a time.

(2) In the base technology, a parameter is determined in a 20 two-step process. That is, in such a manner that a point at which $C_f^{(m,s)}$ takes the minima is detected after a mapping such that the value of the combined evaluation equation becomes minimum is determined. However, instead of this two-step processing, a parameter may be effectively determined, as the

case may be, in a manner such that the minimum value of a combined evaluation equation becomes minimum. In this case, $\alpha E_0 + \beta E_1$, for example, may be used as the combined evaluation equation, where $\alpha + \beta = 1$ may be imposed as a constraint so as to 5 equally treat each evaluation equation. The automatic determination of a parameter is effective when determining the parameter such that the energy becomes minimum.

(3) In the base technology, four types of submappings related to four types of critical points are generated at each 10 level of resolution. However, one, two, or three types among the four types may be selectively used. For instance, if there exists only one bright point in an image, generation of hierarchical images based solely on $f^{(m,3)}$ related to a maxima point can be effective to a certain degree. In this case, no 15 other submapping is necessary at the same level, thus the amount of computation relative on s is effectively reduced.

(4) In the base technology, as the level of resolution of an image advances by one through a critical point filter, the number of pixels becomes $1/4$. However, it is possible to 20 suppose that one block consists of 3×3 pixels and critical points are searched in this 3×3 block, then the number of pixels will be $1/9$ as the level advances by one.

(5) In the base technology, if the source and the destination images are color images, they would generally

first be converted to monochrome images, and the mappings then computed. The source color images may then be transformed by using the mappings thus obtained. However, as an alternate method, the submappings may be computed regarding each RGB 5 component.

Preferred Embodiments for Multivariate Space Processing

A technique for multivariate space processing utilizing the above-described base technology will now be described. In 10 the following, the term "variate" or "variable" is used to describe each of a plurality of variable quantities in a multivariate space. It is to be noted that a variate may also sometimes be called a parameter and these terms may generally be treated in the same manner as the term "dimension", as the 15 case may be. Further, when a variate is referred to as a variate of an object, the variate may also be defined as an attribute of the object.

Fig. 18 shows a structure of a multivariate space processing apparatus 10 according to an embodiment of the 20 invention. This apparatus includes: a preprocessing unit 12 which acquires multivariate data, which in a particular case is object data (OD), and degenerates the object data into three variates; a conversion unit 14 which receives the degenerated data and then generates a first image and a second

image from the degenerated data; a matching processor 16 which performs a matching computation on the first image I1 and the second image I2; a corresponding point file storage unit 18 which stores a corresponding point file F obtained as a result of the matching computation; an intermediate image generator 20 which generates an intermediate image based on the first image I1, second image I2 and corresponding point file F; and a display control unit 22 which performs any processing necessary for displaying the intermediate image thus generated. The conversion unit 14 also generates an authentic intermediate image AIF at a predetermined point between the first image I1 and the second image I2, and outputs this authentic intermediate image AIF to a comparator 24. The intermediate image generator 20 also outputs a virtual intermediate image VIF generated from the first image I1 and the second image I2 to the comparator 24. The comparator 24 compares these two intermediate images, and stores a comparison result in a comparison result file storage unit 26.

Preferably, the preprocessing unit 12 also stores the three variates selected by a user in a variate-selected memory unit 28 for use in subsequent processing. Moreover, the preprocessing unit 12 also preferably has access to comparison results for previous data by referring to the comparison result file storage unit 26. In this way, the preprocessing

unit 12 can display to a user previously selected variates and an indication of which previously selected variates provided a good comparison result. The user can utilize this information to aid in selection of the three variates to be used for the 5 current process.

Fig. 19 is a flowchart showing an overview of the processing in the multivariate space processing apparatus 10. The object data OD are first input to the preprocessing unit 12 where the input object data OD are degenerated into three 10 variates (S300). Preferably, the preprocessing unit 12 displays previously selected variates and some indication of the order of approximation between the authentic intermediate images AIF and the virtual intermediate images VIF for the previously selected variates, as described above. This may 15 assist the user in selecting a certain variate or variates to be targeted in the visualization. This information is particularly effective in a case where there are many variates of the same or similar level of importance.

Moreover, the preprocessing unit 12 preferably displays 20 to the user previously selected variates for each type of object data OD. Thereby, the variates that the user was interested in in the past or the variates that the user chose for a predetermined purpose can be provided to the user to allow easier and more efficient selection of variates by the

user.

Degeneration by the preprocessing unit 12 can be realized by various methods. As a simple example, the three variates the user selected, as they are, are selected and the 5 remaining variates are ignored. As another method, in the course of reducing the variates, a conversion process can be performed on one or more remaining variates at an arbitrary timing. In a case where such a conversion is performed, the converted variates may also be stored in the variate-selected 10 memory unit 28.

Similar to a case where the degeneration method involves a certain dimension being simply cut off, an alternative may be to assign a specific value for that dimension. For example, in a case where a three-dimensional space is 15 degenerated into a two-dimensional space, the height may be set to some constant. Here, it suffices that the degeneration is such that n dimensions are degenerated (reduced) to a dimension number which is less than n. For example, a three-dimensional space may be converted to an arbitrary plane 20 $ax+by+cz+d=0$ or the three-dimensional space may be converted to a curved surface such as $ax^2+by+c=0$. In any event, reducing the number of variates is called degeneration.

After the object data OD have been degenerated into three variates, the degenerated data are input to the

conversion unit 14. The conversion unit 14 then waits for the user to input which variate is to be regarded as a reference among the three variates. The determination of which variate is to be used as a reference may also be set as a default or 5 the like. Once a variate which will serve as the reference (simply referred to as a reference variate hereinafter) is specified, the conversion unit 14 generates the first image I1 and the second image I2 by setting two predetermined values to this reference variate (S302). Again, the predetermined values 10 may be set by the user, may be default values, or may be otherwise determined.

As an example, the reference variate may be time t and the remaining two variates may be x and y spatial coordinates. By examining these variates, a first image at time 15 $t=t_1$ and a second image at time $t=t_2$ can be defined. If an image matching is computed between these two images, an approximate behavior of the two variates x and y in the interval $t=[t_1, t_2]$ can be determined. Thus, it suffices to compute a matching between the two-dimensional images and the 20 computational load can be reduced by selecting a suitable algorithm. The "base technology" provides a suitable algorithm and contributes to producing a suitable visualization.

The first image I1 and the second image I2 thus

generated are output to the matching processor 16. The matching processor 16 performs a matching computation where, for example, the computation may focus on critical points, as described in the base technology (S304). The corresponding 5 point file F generated as a result of the matching computation is stored in the corresponding point file storage unit 18.

The intermediate image generator 20 receives the first image I1, the second image I2, and the corresponding point file F and generates a virtual intermediate image VIF by 10 performing an interpolation computation based thereon. An example of a possible interpolation computation is described in more detail in the base technology. The intermediate image thus generated is output to the display control unit 22. The display control unit performs any necessary processing of the 15 intermediate image and the processed intermediate image is output to a display device (5308). The virtual intermediate image VIF thus generated is also output to the comparator 24 S308).

Fig. 20 shows more detail of S300 shown in Fig. 19. The 20 preprocessing unit 12 first refers to the comparison result file storage unit 26 and the variate-selected memory unit 28 to determine whether or not reference information exists (S300A). Here, "reference information" is used as a generic term for previous comparison results and previously selected

variates. If any reference information exists (Y in S300A), it is displayed (S300B) whereas, if reference information does not exist (N in S300A), the display of the reference information (S300B) is skipped and the process waits for the 5 user to specify selected variates (S300C).

After the user selects three variates by either confirming the displayed reference information or by specifying new variates (Y in S300C), the preprocessing unit 12 degenerates the object data OD into the selected three 10 variates.

We now consider a stock price as an example of how a complicated multivariate object can be converted into a simple model by selecting three variates. As will be understood, there are generally a very large number of parameters that go 15 into determining a stock price. Initially, three variates are tentatively selected based on a rule of thumb or an arbitrary selection after a search algorithm which has enumerated a number of possible variates. Next, a first image and a second image are generated and a virtual intermediate image thereof 20 is generated by matching and interpolation. Thereafter, this virtual intermediate image and the actual stock price (which represents the authentic intermediate image) are compared to determine the suitability of the selection of the three variates. Thus, if the virtual intermediate image is close to

the authentic intermediate image, it is concluded that the three variates have high importance. This can be used or further tested by determining if it is possible to analyze or predict the stock price at another time by using these three 5 variates. If the virtual intermediate image and the authentic intermediate image are not close or if further testing is needed, the virtual intermediate image and the authentic intermediate image may continue to be compared while the three variates are being changed, the selection of the reference 10 variate is being changed and/or the value(s) set for the reference variate is/are being changed.

Fig. 21 shows more detail of S302 shown in Fig. 19. The conversion unit 14, which has received the three variates, first inquires of the user as to which variate is to be 15 regarded as the reference variate (denoted as t hereinafter) (S302A). After the user specifies the reference variate at S302A, the conversion unit 14 requests two specific values of the reference variate t from the user (S302B). These two values are necessary for generating the first image 20 I1 and the second image I2. As noted above, the two values may alternatively be predetermined or calculated values. After the conversion unit 14 acquires the user specified two values $t=t1$ and $t=t2$ at S302B, the conversion unit 14 determines the first image I1 and second image I2 (S302C). Since values of

the remaining two variates may be determined by substituting specific values into the reference variate t , the first image I_1 and the second image I_2 represented by these two variates can be determined efficiently.

5 The processes performed by the matching processor 16 at S304 of Fig. 19 and the intermediate image generator 20 at S306 of Fig. 19 are generally as described in the base technology above.

Fig. 22 shows more detail of S308 shown in Fig. 19.

10 After the intermediate image has been generated by the intermediate image generator 20 (S306 of Fig. 19), the display control unit 22 requires this intermediate image and may also receive the first image I_1 and second image I_2 depending on the application. The display control unit 22 then converts 15 the intermediate image into a data format which can be displayed on the display unit (S308A). Thereby, a two-dimensional image is realized for visualization.

20 The intermediate image is also sent to the comparator 24 by the intermediate image generator 20 as a virtual intermediate image VIF and the comparator 24 compares the authentic intermediate image AIF and the virtual intermediate image VIF (S308B). For example, the comparator 24 may compare the authentic intermediate image AIF and virtual intermediate image VIF at a midpoint $t=(t_1+t_2)/2$ of $t=t_1$ and $t=t_2$. In this

case the conversion unit 14 may be arranged to substitute $t=(t_1+t_2)/2$ as the reference variate and determine the remaining two variates at that value so that a two-dimensional image can be generated. This image is generated based on the 5 actual data and is, in this sense, an authentic intermediate image. Similarly, the intermediate image generator 20 may use the same reference variate value $t=(t_1+t_2)/2$ so that a virtual intermediate image at that reference variate value is generated through an interpolation computation.

10 The comparator 24 takes a difference of the authentic intermediate image AIF and the virtual intermediate image and determines whether the result is desirable or not according to an amount of the difference. The comparison result is stored in the comparison result file storage unit 26 (S308C).

15 Alternatively, the comparator 24 may simply compute the difference and may store this computed difference in the comparison result file storage unit 26. In this case, the preprocessing unit 12 may determine whether or not the comparison result is desirable.

20 Figs. 23, 24 and 25 show experimental visualization results according to the present embodiment. Here, data on atmospheric pressure in the vicinity of the center of a tornado are considered as the object data. Fig. 23 and Fig. 24 show distributions of atmospheric pressure viewed from a

transverse direction at predetermined times $t=t_1$ and $t=t_2$, respectively. Thus, Fig. 23 may be considered the first image I_1 and Fig. 24 may be considered the second image I_2 described above. Fig. 25 shows a virtual intermediate image at

5 $t=(t_1+t_2)/2$ between the first image I_1 and the second image I_2 . Now, suppose that the z axis is placed in the longitudinal direction of the tornado, then the first image I_1 and the second image I_2 , shown in Fig. 23 and Fig. 24, respectively, can be thought of as three-variate made up of

10 (1) two dimensional information acquired when the object data are projected on a plane parallel to the z axis and (2) the time. Thus, the processing of the preprocessing unit 12 in this example is one in which the object data are projected on a plane parallel to the z axis and the time parameters are

15 stored.

Further, the processing of the conversion unit 14 can be considered as selecting time t as the reference variate and obtaining two dimensional images at the times t_1 and t_2 . As can be seen from the result shown in Fig. 25, visualization

20 results can be obtained efficiently by utilizing a simplified method according to the present embodiment.

In other words, the method and apparatus for multivariate space processing according to the above embodiment are implemented in the tornado example such that

(1) the first image and the second image are acquired by projecting three-dimensional data of an object on a predetermined x-y plane and (2) a matching between the first image and the second image thus acquired is computed so that

5 intermediate states in the range of time $t=[t_1, t_2]$ can be visualized by interpolating the two-dimensional images at time $t=t_1$ and time $t=t_2$. Computationally, once the matching is completed, there will not be much time needed for the interpolation. Moreover, since the interpolation can be

10 performed at any point by arbitrarily varying the time t , the computational load becomes extremely low compared to the conventional visualization where a three-dimensional computation is performed sequentially. Thus, the present embodiment is effective particularly when one wants to quickly

15 visualize the result showing a simulation in as simplified and convenient a manner as possible.

As is seen in the tornado example, there is an implicit relationship between degeneration and projection of object data. That is, when reducing or degenerating the parameters

20 of a multidimensional phenomenon or object, the object may be "viewed" from various "angles", an inspection may be carried out to determine an angle at which a most characteristic image is obtained, and the degeneration to view the object at that particular angle may be repeated. Using a simple example,

consider a case where a three-dimensional object is projected as a "shadow picture" on a plane. For most three-dimensional objects, there will be an angle at which the original object can be relatively easily grasped from the shadow picture. A

5 projection operation using this angle corresponds to the degeneration.

Though the present invention has been described based on the above embodiments, the present invention is not limited to these specific embodiments alone and various modifications thereto are also effective as encompassed by the present 10 invention. Some of such the modifications will be described here.

In the present embodiment, the tornado example was used to illustrate a phenomenon in three-dimensional space with the 15 added dimension of time. However, the present invention is not limited to three or four dimensions and can also be applied to handle multivariate phenomena or objects. For example, in analyzing a complex economic phenomenon, a similar visualization process can be performed in a simple manner by 20 initially specifying three variates that the user should pay attention to among various parameters. Then, by varying selection of the three variates, it is possible to determine which particular variates have a higher degree of importance in the economic phenomenon. Moreover, based on the comparison

results of the comparator 24, the variate or variates which are important in the economic phenomenon can be more easily determined on a case-by-case basis.

In other words, by implementing the multivariate space processing method according to the present invention, it is possible to avoid the complex modeling which is generally required in conventional simulation and analysis of complicated multivariate objects or phenomena, and visualization can be realized using a simple and convenient method. Thus, one of the main advantages of the embodiments of the invention lies in the proposition and realization of this new methodology by which arbitrary phenomena can be understood by a process of relatively simple matching of two-dimensional images and without modeling.

Although the present invention has been described by way of exemplary embodiments, it should be understood that many changes and substitutions may be made by those skilled in the art without departing from the spirit and the scope of the present invention which is defined by the appended claims.